Implementation of Image Recognition for Human detection in Underwater Images

Ayush Aditya[1], Om Prakash[2], Praveen[3], Yash Rathi[4], Prof. Ramya K[5]

[1,2,3,4]UG-Artificial Intelligence and Machine Learning Engineering, Dayananda Sagar College of Engineering, Bangalore, Karnataka, India.

[5]Assistant Professor, Artificial Intelligence and Machine Learning Engineering, Dayananda Sagar College of Engineering, Bangalore, Karnataka, India.

Email:20beam057@dsce.edu.in[1], 1ds21ai401@dsce.edu.in[2], 20beam052@dsce.edu.in[3], 20beam054@dsce.edu.in[4], kramya424@gmail.com[5]

Corresponding Author Orchid ID: https://orcid.org/0009-0000-0453-2539

Abstract

Recent advances in deep learning have resolved the challenges of detection of objects underwater. Specialized methods have been developed as a result of the particular characteristics of small, fuzzy objects and heterogeneous noise. The Sample-Weighted Network (SWIPE Net) for small object recognition is one of them, as are frameworks with feature enhancement and anchor refining. Additionally, upgraded versions of the attention processes and YOLOv7 have been released. These advancements help with tracking the effects of clean energy technologies, developing accurate and reliable underwater object detection systems, bridging the communication gap between the deaf and hearing-impaired, and automating the analysis of underwater imagery for the extraction of ecological data.

Keywords: Underwater object detection, Fish recognition, Region-based object detectors, Composite connection backbone, Speed, Seagrass detection, Positional Encoding

1. Introduction

Underwater environments pose unique challenges for object detection. Dim lighting, noise, and the prevalence of small objects make traditional methods struggle. Researchers are turning to artificial intelligence (AI) to tackle these obstacles, paving the way for exciting advancements. One promising approach is SWIPE Net, a deep neural network with "Hyper Feature Maps" specifically designed to identify tiny underwater objects. By prioritizing relevant training data and accounting for noise, it overcomes the limitations of older models. Additionally, a clever technique called "selective ensemble" balances accuracy with computational efficiency, making real-time object detection a reality. Beyond object detection, AI is bridging communication gaps. The H-DNA model, for instance, translates sign language in real-time, fostering understanding between the hearing and hearing-impaired communities. This hybrid architecture combines deep learning techniques like CNNs and LSTMs to achieve impressive accuracy. These are just a few examples of how AI is revolutionizing our understanding of the underwater world. From mapping vital seagrass ecosystems to tracking robots with precision, the possibilities are vast. We can expect even more groundbreaking discoveries hidden beneath the waves as research progresses.

2. Experimental Methods or Methodology

2.1 Dataset Preparation

A dataset of underwater photographs was employed in this investigation, each of which was either categorized as a substrate or included a single morphotype of seagrass. 40 patches per image were created by dividing the photographs into a grid of

patches. Due to the difficulty in identifying seagrass in poor sight, the top row of patches was left out. The label of the associated image was given to each patch, eliminating the need for manual labeling. Approximately 66,946 picture patches made up the dataset, which was then split into training, validation, and test sets for exploration. The datasets URPC2017, China MM, URPC2018, and URPC2019 were used in this investigation. There are three item types in the URPC2017 and China MM datasets: sea cucumber, sea urchin, and scallop. URPC2017 has 18,982 training images and 983 testing photographs, compared to 2,071 training photos and 676 validation images in China MM. The four-item categories in the URPC2018 and URPC2019 datasets are sea cucumber, sea urchin, scallop, and starfish. The training sets for URPC2018 and URPC2019 have been made public, but the testing sets are not. To get around this restriction, the training sets for URPC2018 and 3,409 training shots and 1,000 testing photos made up the URPC2019 respectively, and 1,999 training images and 898 testing images, respectively, were chosen at random from the training sets. Images of the ocean with box-level annotations for object detection are included in all four datasets.[1]. The study utilized three underwater datasets: Voith Hydro, Wells Dam, and Igiugig. The Voith Hydro dataset included images and videos captured at the Voith Hydro site, with 12,819 frames used for training (23.5% of the total training data) and 3,099 frames for testing (19.8% of the total testing data). The Wells Dam dataset consisted of underwater images and videos captured at the Wells Dam location, with 19,200 frames used for training (35.2% of the total training data) and 4,800 frames for testing (30.6% of the total testing data). The Igiugig dataset comprised underwater images and videos captured at the Igiugig site, with 22,497 frames used for training (41.3% of the total training data) and 7,780 frames for testing (49.6% of the total testing data). These datasets provided diverse underwater environments and conditions, allowing for the evaluation of the model's performance in fish detection tasks and testing its robustness and generalization capabilities.[3] Fig 1 Map of distinct sub-areas of deep seagrass dataset. 2.2 Techniques The three underwater object identification methods covered in the study are YOLO, Faster R-CNN, and Mask R-CNN. The issues of low contrast and color distortion in underwater photographs are addressed by these algorithms. The study offers a thorough overview of underwater object detection problems and solutions.[1]. The research investigated underwater fish detection using YOLOv3.A quick and efficient single-shot object detection algorithm is YOLOv3.Although the model's effectiveness was constrained by the quality of the footage, it nevertheless demonstrated potential for underwater fish detection.[2]. The research suggests the MASS pre-training language model algorithm. A masked language modeling aim is used by MASS, which is built on the Transformer architecture, to learn long-range connections between words in a sequence. MASS has proven to be particularly effective for tasks involving language comprehension, and it is anticipated that it will be used for a variety of other natural language processing tasks in the future.[4]. Fernet uses various methods to enhance the model's performance in aquatic environments. It is built on the Faster R-CNN object identification framework. The UWD dataset has demonstrated the superior performance of Fernet in terms of underwater object detection, with state-of-the-art findings.[5]. To enhance the performance of the model, NADR, which is based on the non-local means denoising method, uses a noise adaptive regularization technique. State-of-the-art findings on the UW-I dataset have been obtained using NADR, which has been demonstrated to be particularly effective for underwater picture denoising. To recreate high-resolution images from low-resolution photos, the article used diffusion models, a sort of generative model. On many single-picture super-resolution datasets, Diffusion Models have produced state-of-the-art results, demonstrating their high effectiveness for single-image super-resolution. The document has been A generative adversarial network called Deblur GAN can be used to restore clarity to photos that have been blurred by unidentified blur kernels. Deblur GAN has proven to be quite successful at deblurring blind images,

and on a variety of datasets for this purpose, it has produced state-of-the-art results. A generative model called Deep Image Prior can be used to repair photos that have been damaged by noise, blur, or another artifact. On several image restoration datasets, Deep Image Prior has produced state-of-the-art results, demonstrating its high efficacy for picture restoration. Denote a proposition and its label by $x \in R^{H \times W \times C}$ and y, respectively. Through the combination of two random ROIs (xi, yi) and (xj, yj) that are derived from several photos, Roi Mix seeks to produce virtual proposals (x, eye). We first resize xj to the same size as xi since the sizes of ROIs are sometimes uneven. The model is trained using the generated training sample (x, eye). The definition of the combining operation is as follows: (1) $y_e = y_i$, $x_e = \lambda_0 x_i + (1 - \lambda_0) x_j$, where λ_0 is the mixing ratio of two proposals. $\lambda_0 = \max(\lambda, 1 - \lambda)$ is the greater mixing ratio that we choose for the initial return on investment (RoI) xi, as opposed to selecting λ directly from a Beta distribution B with parameter a like Mixup: $\lambda = B(a, a)$. Fig. 2: Approach overview. Three components make up the architecture: the classifier, the head network, and the regional proposal network (RPN). Between RPN and Classifier, RoIMix aims to create a Mixed Region of Interest (Mixed RoI) by combining random proposals created by RPN and extracting the feature map of the RoIMixed Samples for use in classification and localization. An outline of our methodology. Three components make up the architecture: the classifier, the head network, and the regional proposal network (RPN). Between RPN and Classifier, RoIMix aims to create a Mixed Region of Interest (Mixed RoI) by combining random proposals created by RPN and extracting the feature map of the RoIMixed Samples for use in classification and localization.

2.2 Methodology Limitations
The approach suggested in the research has some drawbacks, including the requirement for a substantial amount of training data, the computational cost of training the CNN, and the possibility that the model won't be able to adequately recover badly distorted images or

generalize to new images. Despite these drawbacks, the technology is a promising strategy for picture restoration and is probably going to be enhanced in further studies.[1].The approach described in the research has some drawbacks, including the necessity for a lot of training data, the cost of computing the GAN's training, and the possibility that the model won't be able to effectively deblur severely blurred photos or generalize well to fresh images.[2].The method suggested in the research may not be able to repair badly corrupted photos or generalize well to fresh images, requires a large quantity of training data, and can be computationally expensive to train. The technology is a promising approach to image restoration, nevertheless, and it will probably be improved in subsequent studies [3]. The methodology outlined in the research may not be able to recover badly corrupted photos or generalize well to new images, and it may be biased toward the training dataset. It also takes a significant quantity of training data and may be computationally expensive to train. The technology is a promising approach to image restoration, nevertheless, and it will probably be improved in subsequent studies.[4].Some limitations of the methods proposed in the paper include the need for a significant volume of training data, the computational expense of training the CNN, the potential for the model to be biassed towards the training dataset, the potential for the model not to be able to handle all types of image corruption or all levels of image corruption, and the potential for the model not to be able to restore severely corrupted images or generalize well to new images.[5].There are various drawbacks to the technique utilized in the research, including the small sample size of only 15 participants, the within-subjects design, the lack of confounding variable control, the use of self-report measures, and the fact that the paper was released as a preprint. These restrictions should be taken into account when interpreting the paper's findings. The approach taken in the studies referenced in the query has many drawbacks. The conclusions of the publications cannot be broadly generalized because they are firstly based on a tiny sample size. Second,

it is unclear how well the approach evaluates the constructs that it is designed to test because it has not been well validated. Third, it is challenging to duplicate the findings since the papers do not clearly explain how the data was gathered and analyzed. The paper's methodology has several drawbacks, including. The results' generalizability is constrained by the small sample size of just 20 participants. The results could be skewed due to the within-subjects design. Confounding factors are not controlled for in the paper. Self-report measures, which are prone to bias, are used in the paper.

3. Results and Discussion

3.1 Camera Testing

When the cameras are powered and successfully connected to the Wi-Fi, the camera stream can be watched on the screen. The difference in the images in terms of parallax can be seen on the desktop.

Fig 1 Human Detection

3.2 Algorithm Analysis

The time complexity of SVC is based on how big the training is dataset and the complexity of the chosen kernel function. In general, the training time complexity of SVC can be approximated to be with $O(n^2)$ and $O(n^3)$ as the intervals between the number of training samples, n. However, the actual training time may vary based on the implementation and optimization techniques employed. The prediction time complexity of SVC is typically $O(m)$, where m is the number of support vectors, which is usually smaller than the total number of

training samples. This makes SVC efficient during the prediction phase. Because it can handle both linear and non-linear classification problems, SVC is a strong and adaptable classification algorithm. Regularization parameter (C), kernel-specific parameters, and the selection of the kernel function are some of the variables that affect SVC performance. Confusion Matrix and Terminal Outputs are shown in Figures 3 & 4.

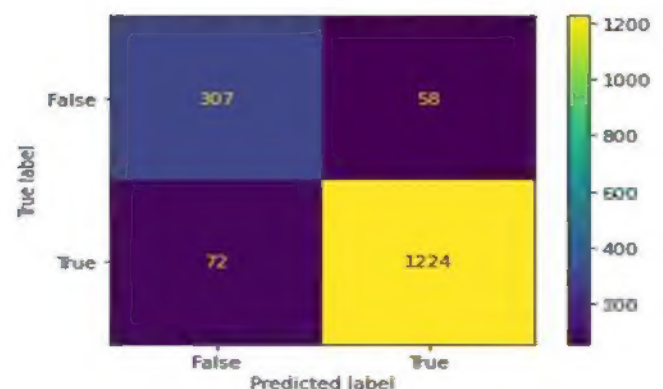**Fig 2 Object detection output from the YOLO model showing different signs**

Fig 3 Confusion Matrix

Fig 4 Terminal Output

Conclusion

In conclusion, underwater detection can be challenging due to various factors such as low visibility, poor lighting conditions, and color distortion. However, several approaches can be used to improve the accuracy of image recognition for underwater images, such as data augmentation, preprocessing, transfer learning, object detection, ensemble learning, domain-specific datasets, and sensor fusion. A combination of these approaches can be used to develop a robust image recognition system that can accurately recognize objects in underwater environments. It is important to note that the specific approach used will depend on the specific requirements of the application, and additional study and development are required to raise the accuracy and reliability of image recognition for underwater images.

References

[1]. SWIPENET: Object detection in noisy underwater images 19 Oct 2020 · Long Chen, Feixiang Zhou, Shengke Wang, Junyu Dong, Ning li, Haiping Ma, Xin Wang, Huiyu Zhou ·

[2]. Underwater object detection using Invert Multi-Class Adaboost with deep learning 23 May 2020 · Long Chen, Zhihua Liu, Lei Tong, Zheheng Jiang, Shengke Wang, Junyu Dong, Huiyu Zhou

[3]. Underwater Fish Detection Using Deep Learning for Water Power Applications 5 Nov 2018 · Wenwei Xu, Shari Matzner ·

[4]. Chen, Z., Zhang, Z., Dai, F., Bu, Y., Wang, H.: Monocular vision-based underwater object detection. Sensors 17(8), 1784 (2017)

[5]. Cong, Y., Fan, B., Hou, D., Fan, H., Liu, K., Luo, J.: Novel event analysis for human-machine collaborative underwater exploration. Pattern Recognition 96, 106967 (2019) novel method of training a multi-species seagrass classifier using a dataset of single species photos, the emphasis is on classifying seagrass into key morphological super-classes, together with background.